

Analyzing 2 Categorical Variables

- You can create categorical variables ...

by grouping quantitative
2 way tables

- Categorical variables are often shown ...

AGE
0-10
10-20
20-30
⋮

Gender

	Frosh	Soph	Junior	Senior	Total
Male					
Female					
Total					

Σ females
 Σ Frosh
 cells
 n

Identify:

- Row variable
- Column variable
- Values of the variable
- Total (n)
- Cells
- Totals

Gender
Grade

2 x 4

- where original data is

Example: Hospitals

	Hospital A	Hospital B
Died	63	16
Survived	2037	784

2100

800

79
2821
2900

- What percent of people died?

$$79/2900 = 2.72\%$$

- Of those people that went to Hospital A, what percent died?

$$\frac{63}{2100}$$

- Of those people who went to Hospital B, what percent died?

$$\frac{16}{800}$$

- Of those people who died, what percent went to Hospital A?

$$\frac{63}{79}$$

- What percent of people died and went to Hospital B?

$$\frac{16}{2900}$$

OR $\frac{(63 + 16 + 784)}{2900}$

2 types of Distributions for Categorical Variables

1) MARGINAL DISTRIBUTIONS

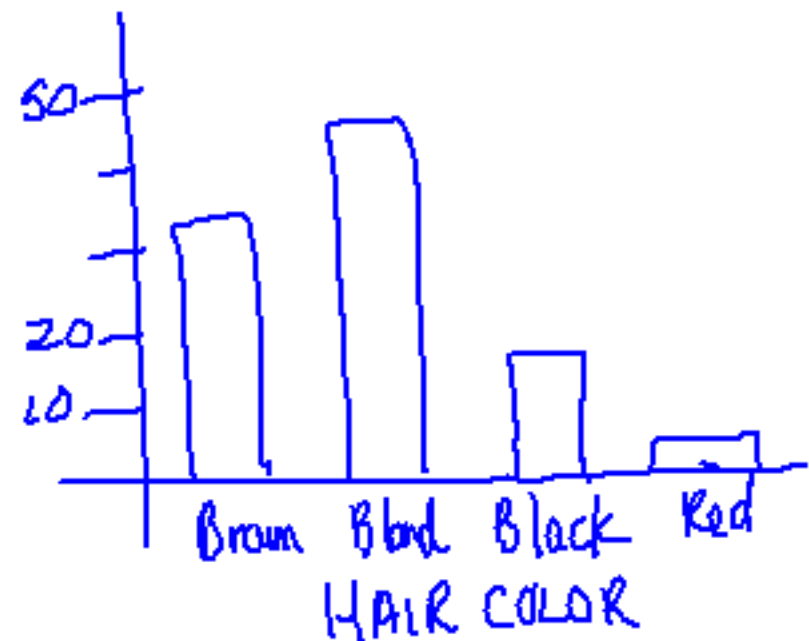
- How to make: $\text{margins (totals)} \div n \text{ (overall total)}$
- Looking for ... overall % of each value of the variable
- ALWAYS ...
in %

- Example: Hair color vs. Gender

	Brown	Blonde	Black	Red	Total
MALE	26	24	10	3	63
FEMALE	20	35	12	6	73
TOTALs	46	59	22	9	136

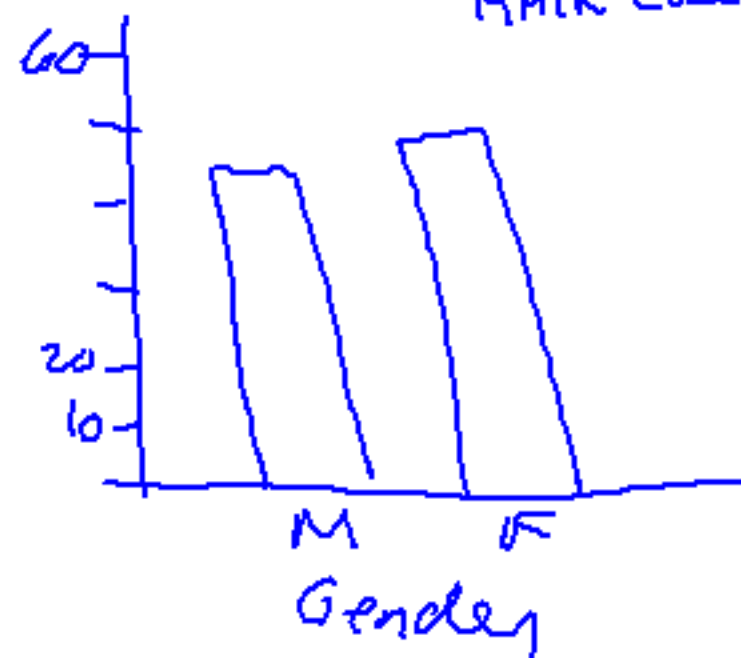
- Find the marginal distribution for the HAIR COLOR variable

Values: Brown $46/136 = 33.82\%$
 Blonde $59/136 = 43.38\%$
 Black $22/136 = 16.18\%$
 Red $9/136 = 6.62\%$



- Find the marginal distribution for the GENDER variable

Values: Male $63/136 = 46.32\%$
 Female $73/136 = 53.68\%$



- Represented Visually: BAR CHART

1) CONDITIONAL DISTRIBUTIONS

- Look at ... *one variable @ a time*

Then look at ... *each value separately*

- Break down ... *each value into its parts (cells)*
- ALWAYS ...
in %

- Example: Hair Color vs. Gender

- Find the conditional Distribution for the HAIR COLOR variable

Values: Brown M-26/46 F-20/46
 Blonde M-24/59 F-35/59
 Black M-10/22 F-12/22
 Red M-3/9 F-6/9

	Brown	Blonde	Black	Red	Total
MALE	26	24	10	3	63
FEMALE	20	35	12	6	73
TOTALs	46	59	22	9	136

- Find the conditional Distribution for the GENDER variable

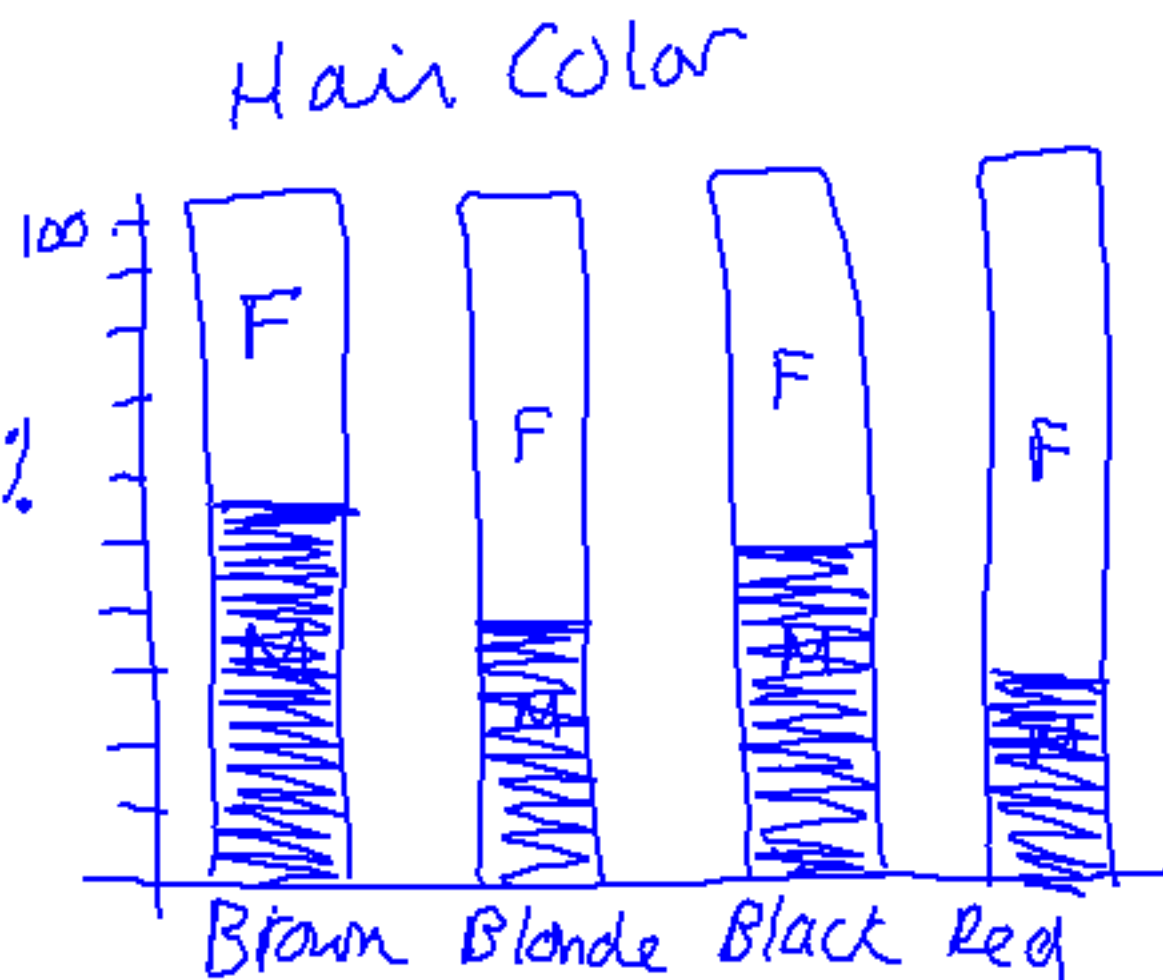
Values: male

Br. 26/63 41.27%
 Blond 24/63 38.1%
 Black 10/63 15.87%
 Red 3/63 4.76%

Female

Brown 20/73 27.4%
 Blond 35/73 47.95%
 Black 12/73 16.44%
 Red 6/73 8.22%

- Represented visually: SEGMENTED (or STACKED) BAR GRAPH
 - Each bar = 100%
 - Values of variable on the x-axis
 - Bars are segmented into parts of each value



▨ = male

